

文章编号: 1671-0576(2023)04-0022-07

基于多尺度链接的红外与可见光图像融合

刘兆丰, 姜家瑞, 傅迎华

(上海理工大学光电信息与计算机工程学院, 上海 200082)

摘 要: 提出了一种红外和可见光图像的融合方法。该方法采用深度学习中的生成对抗网络(generative adversarial network, GAN)来融合两种模态图像,融合过程主要通过网络架构中生成器和鉴别器之间的对抗实现。生成器采用多尺度链接架构,有效提取并利用源图像的深层与浅层特征信息。同时鉴别器采用了与传统全局鉴别器不同的局部鉴别器,确保融合后的图像充分包含源图像的信息与特征分布。经实验验证,采用该方法融合后的图像能有效包含两种源图像各自的特征。

关键词: 红外图像; 可见光图像; 生成对抗网络; 图像融合

中图分类号: TP391.41

文献标志码: A

DOI: 10.3969/j.issn.1671-0576.2023.04.004

Infrared and Visible Image Fusion Based on Multi-scale Link

LIU Zhaofeng, JIANG Jiarui, FU Yinghua

(School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200082, China)

Abstract: A fusion method for infrared and visible image was presented. The method used generative adversarial network (GAN) of deep learning to fuse two modal images. The fusion process was mainly achieved through adversarial interactions between generators and discriminators of the network architecture. The generator employed a multi-scale link architecture to allow effective extraction and utilization of deep and shallow-level features from the source images. Moreover, the local discriminator which was distinct from traditional global discriminator was used to ensure comprehensive incorporation of the information and feature distributions from the source images in the fused output. Experimental results demonstrate the effectiveness of the proposed method in preserving the distinctive characteristics of both source images in the fused output.

Key words: infrared image; visible image; GAN; image fusion

0 引言

图像融合^[1]是指对不同传感器所拍摄的同目标的图像数据进行一系列预处理操作,从源图像数据中分别提取最有意义的信息加以综合,最终生成新的对后续应用有利的单一图像的过程。通常情况下,当目标与背景具有相同的颜色或特征时,单一传感器无法获取能够准确区分二者的图像。可见光图像分辨率较高,包含了更多的纹理细节,但比较依赖天气和光照条件。当天气条件恶劣或者夜晚光线微弱时,传感器所采集的图像无法很好地显示目标信息。红外传感器成像则不依赖于天气条件,虽所获取的图像往往分辨率较低,但其能够利用目标的热辐射信息有效突出目标。例如,2020年疫情开始爆发,为了有效控制疾病的蔓延,人们在各地的交通枢纽上都安装了红外体温检测装置,这些装置对疾病的初步筛查以及遏制疾病扩散起到了至关重要的作用^[2]。

红外图像与可见光图像的不同特性使得这两种图像的融合技术在民用和军事领域都得到了广泛的应用。例如:在农业领域中,通过无人机在空中获取的地面图像,对农作物进行病虫害监测;在环境治理中,通过图像对水域污染进行监测;在医学领域中,将采用磁共振成像(MRI)与正电子发射断层扫描术(PET)两种手段获取的影像进行融合,辅助医生对病理进行分析^[3]。根据目标信息实现对目标的检测与跟踪一直以来都是红外制导领域的核心技术,并已经成为各国军事领域的主要研究热点之一。

由于传感器技术的限制,单个传感器无法捕获全部的场景信息,而多模态图像融合可将多个传感器获取的图像中的重要信息进行融合,实现对场景的有效描述。红外与可见光图像融合便是最具代表性的多模态图像融合任务之一。早期的红外与可见光图像融合技术采用的方法有基于多尺度变换的方法^[4-5]、基于稀疏表示的方法^[6-7]、基于显著性的方法^[8]等。早期融合方法采用人工设计融合规则,由于场景复杂,依赖于经验的人工规则方法融合策略过于粗糙,没有考虑源图像的特征差异及多样性,融合效果非常有限^[9]。

近年来随着深度学习技术的快速发展,基于深度学习的图像融合方法逐步取代了传统的融合方法。将深度学习技术引入到图像融合中,通过设置合适的损失函数自适应地改进图像特征融合的策略,可以弥补传统图像融合方法的缺陷。

LI等^[10]提出了一种端到端的残差融合网络结构(end-to-end residual fusion network-nest, RFN-Nest)用于红外和可见光图像的融合,并采用细节保持损失函数和特征增强损失函数来训练RFN-Nest。LI等^[11]提出了一种基于编解码器的网络架构,利用编码网络提取图像特征,再通过解码网络得到融合图像。该方法有效利用了图像深层特征,采用了两个融合层(融合策略)来融合图像特征。MA等^[12]首次将生成对抗网络(generative adversarial network, GAN)引入图像融合,提出了基于GAN的图像融合(FusionGAN)方法。生成器生成包含红外强度和可见光梯度信息的融合图像,鉴别器迫使融合图像包含更多源图像的细节特征,二者使得融合图像同时包含红外图像的热辐射特征和可见光图像的纹理特征。

基于GAN在图像融合中的有效性,本文提出一种基于GAN的红外与可见光图像融合架构,该架构可避免传统方法需要人工设计融合层的弊端。生成器采用多尺度链接架构,以便更好地捕捉源图像中的特征差异与全局结构,同时将红外图像的前景信息用于图像融合,使生成器生成更具丰富性和真实感的融合图像。鉴别器采用具有局部感知能力的局部鉴别器,局部鉴别器以图像局部区域为单元进行判别,能够更好地感知和理解图像的局部信息,更准确地评估图像的真实性。

1 深度融合技术发展

1.1 深度学习技术

深度学习于20世纪80年代被提出,是机器学习的重要分支之一。人工神经网络是深度学习的前身,相当于深度学习的早期模型。但该网络由于采用单层感知机,只能处理线性问题而无法解决非线性问题,这导致其后续发展出现了短暂的停滞。

2006年 Geoffrey 等提出了深度信念网络概念,通过预训练可以使网络的参数接近最优解,再通过一定的微调即可实现网络优化。该方法使网络的运算速度得到了大幅提升^[13]。此后,深度学习技术迎来了爆发式发展,各种结构的神经网络被提出,卷积神经网络(CNN)便是其中的代表性模型之一。后续随着计算机算力的提升,深度学习模型的特征提取时间大大缩短,图像特征处理的精度也得到了提高。

2014年 GAN 的提出是深度学习的又一突破性进展,该网络将生成模型和鉴别模型紧密联系起来。将 GAN 应用到图像融合领域,可在大幅提高融合图像效果的同时,缩短图像融合时间^[14]。2015年基于编解码结构的深度学习网络 U-Net 在 MICCAI 会议上被提出用于生物医学图像分割。U-Net 在较少训练样本情况下具有较强的特征提取能力和分割性能,该特点使得其被广泛应用到更多领域。

1.2 生成对抗网络

GAN 主要由生成器和鉴别器两个模块组成,其结构如图 1 所示。

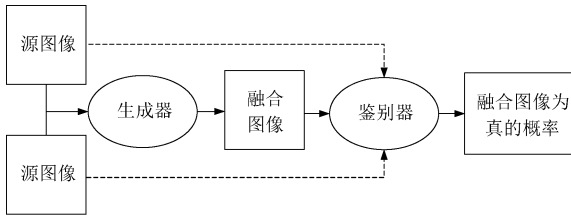


图 1 生成对抗网络结构框图

生成器将接收的红外图像与可见光图像作为源图像输入,从源图像中分别提取特征并生成融合图像。鉴别器接收融合图像后,判断该融合图像是否为真实图像。生成器的主要任务是生成足够真实的融合图像,而鉴别器的任务则是尽可能地将融合图像识别为假图像。由于这两个模块的相互对抗,生成器生成的融合图像将尽可能包含可见光图像和红外图像的特征信息,而鉴别器则将融合图像的纹理信息与可见光图像的纹理信息进行对比,同时将融合图像的热辐射信息与红外图像的热辐射信息进行对比,通过信息对比的结果来判断融合图像的真实性。

在经过一段时间训练后,生成器生成的融合

图像包含了可见光图像和红外图像的主要特征,而鉴别器无法区分该融合图像是真实的源图像还是融合后的假图像。这时模型结构已经达到最优,鉴别器对输入图像判断为真或假的概率均为 50%。

1.3 U-Net

U-Net 是一种具有编码-解码器结构的深度学习网络,其整体架构如图 2^[15] 所示。

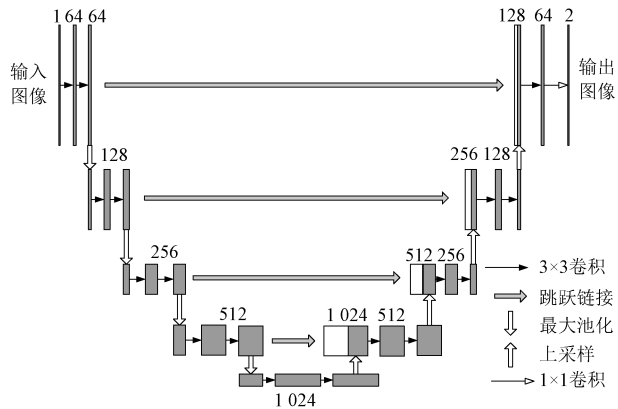


图 2 U-Net 架构

U 型结构的左半部分为编码器,向右的实心箭头代表卷积操作,向下的白色箭头代表池化操作。输入图像经过逐级下采样后获得不同分辨率下的特征图,这些特征图分别代表图像的由低层特征到高层特征的一系列特征表示。U 型结构的右半部分为解码器,由一系列与下采样相对应的上采样构成,上采样如图 2 中向上的白色箭头所示。上采样操作主要包括反卷积和线性插值等。跳跃链接主要用于融合相同特征层的编码器与解码器特征,如图 2 中的灰色箭头所示。上采样的主要目的是利用所获得的一系列编码特征恢复图像目标的概率图。该网络通过设定不同层的损失函数获得图像目标精确的特征表示,从而准确地分割目标。

U-Net 改变了传统深度学习网络的结构,带权重的交叉熵损失函数的构建使得该网络能够在较小的数据集上获得很好的性能。该网络不需要大量带标注的训练样本,仅通过对图像进行多层特征提取,即可提高数据增强的使用深度,有效地捕获图像的深层信息与浅层信息。

2 基于多尺度链接的图像融合

2.1 网络架构

本文采用基于 GAN 框架的多尺度链接方法生成融合图像,其图像融合的网络架构如图 3 所示。生成器采用基于 U-Net 的多尺度链接架构,对图像的深层特征与浅层特征进行有效融合。鉴别器采用与传统全局鉴别器不同的局部鉴别器,能使生成器在生成融合图像的过程中更加注重图像细节。整个网络架构中各包含一个红外热辐射信息鉴别器 Dir 和可见光纹理信息鉴别器 Dvi,以及一个生成器。生成器从两幅源图像即红外图像(IRI)与可见光图像(VI)中提取各自特征,并生成融合图像,而两个鉴别器则负责从输入图像中正确区分真实图像与假图像。

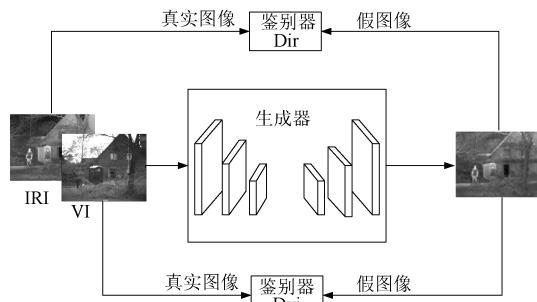


图 3 图像融合网络架构

在不断地对抗学习的过程中,生成器为了欺骗鉴别器将迫使生成结果包含更多的源图像特征,鉴别器则不断提升自身辨别图像真假的能力。为了进一步提升生成器和鉴别器各自的能力,采用多尺度链接架构和局部鉴别器,让生成器与鉴别器更加注重对深层以及局部特征的应用。

2.2 多尺度链接架构

传统的 U-Net 在图像分割任务中取得了很好的效果,但该网络仍然存在单阶段的上下文信息提取能力不足的问题。借鉴文献[16],本文设计了一种基于编码-解码器的多尺度跳跃链接架构,如图 4 所示。

该架构基于 U-Net,将红外图像与可见光图像在通道上进行拼接,并将其输入到编码器网络中进行多尺度的特征提取。编码器网络包括四个编码器,每个编码器独立进行卷积、批量归一化

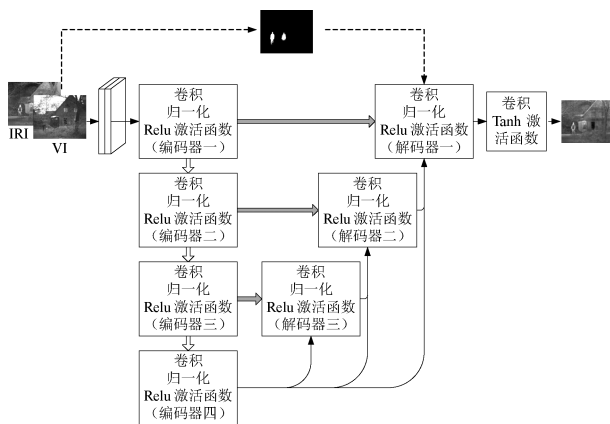


图 4 多尺度跳跃链接架构

(batch normalization, BN) 和 ReLU 激活函数运算,可以提取并有效利用图像深层特征。同时跳跃链接可以将浅层高对比度的特征直接用于最终图像生成,使生成的图像能很好地保留源图像的显著特征。图 4 中白色箭头代表下采样操作,灰色箭头代表跳跃链接,黑色实线箭头代表上采样操作。将深层抽象的信息直接用于浅层解码器的图像生成,可以促进对图像深层信息的利用。其次,通过解码器进行图像恢复。最后,利用预训练的 R^3net ^[17],计算红外图像热目标的掩膜,将源图像中突出的目标图像传至解码器用于融合图像生成,以进一步突出源图像的红外特征信息。

2.3 局部鉴别器

在生成对抗网络中,传统的全局鉴别器最终输出的是一个标量值,并基于整幅图像的特征来判定输入图像的真或假。但传统的全局鉴别器对图像局部细节的真实概率考虑得并不充分,对于一些分辨率高、细节丰富的图像的鉴别效果并不好。因此本文引入了局部鉴别器。图像经过多层卷积后,将输入源图像映射为一个 $N \times N$ 的概率矩阵,矩阵中的每个元素值代表了源图像中的一小块区域为真实图像的概率。对该概率矩阵求平均,将得到的均值作为局部鉴别器的最终输出。这样可以确保在对抗过程中,生成器在生成图像时更加注重图像细节内容的保留。局部鉴别器工作原理示意如图 5 所示。局部鉴别器最后输出的不是一个标量值,而是一个 $N \times N$ 的概率矩阵,其中 D_1, D_2, D_3 表示源图像中对应的块为真的概率。采用这样的方法对每个块进行差别判别,

即可实现对局部图像特征的辨别。通过对得到的概率矩阵求均值,能够实现相比单标量输出的全局鉴别器更为精确的整体差异表示,相当于对整体进行加权求和平均,能够实现比全局鉴别器更为合理的损失表示。

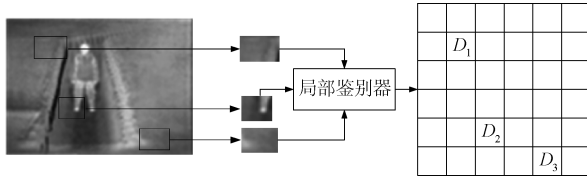


图 5 局部鉴别器工作原理示意图

3 实验验证

3.1 评价指标

红外与可见光图像融合效果可以分别从定性与定量两个方面来评价。定性评价主要从融合图像的视觉效果上进行比较。由于红外图像与可见光图像的融合效果目前尚没有统一的定量评价指标,因此选取多种指标来进行定量比较。

图像的熵(entropy, EN)^[18]是基于信息论定义的,用来衡量融合图像所包含的信息量。数学上,信息熵 H 定义为

$$H = -\sum_{l=0}^L p_l \log_2 p_l \quad (1)$$

式中: L 为灰度级数; p_l 为融合图像中灰度级 l

出现的概率,可由归一化直方图获得。熵越大,融合图像包含的信息越多,融合方法的性能越好。

标准差(standard deviation, SD)^[19]反映了图像中单个像素的灰度值与平均灰度值的距离。在数学上,标准差 S 定义为

$$S = \sqrt{\sum_{i=1}^M \sum_{j=1}^N (F(i, j) - \mu)^2} \quad (2)$$

式中: $F(i, j)$ 为尺寸为 $M \times N$ 像素的融合图像的灰度函数,其中 i, j 分别为像素坐标; μ 为融合图像的灰度均值。高对比度的区域总是会吸引人们的注意力,而高对比度的融合图像往往会产生更大的灰度值的标准差,这意味着融合后的图像会获得了更好的视觉效果。

3.2 实验结果

在 TNO 多波段图像数据集^[20]上进行训练与验证。TNO 数据集目前由三个包含不同军事和监视场景的独立图像集组成,分别为增强视觉(波长为 390 nm~700 nm)图像集、近红外(波长为 700 nm~1 000 nm)图像集和长波红外(波长为 8 nm~12 nm)夜间图像集,图像集中包含了不同背景(如农村、城市)下的不同物体和目标(如人、车辆)图像。

将本文所提融合方法(Ours)与 RFN-Nest、FusionGAN 两种经典融合方法获得的图像进行融合性能定性比较,源图像及融合图像的定性比较如图 6 所示。

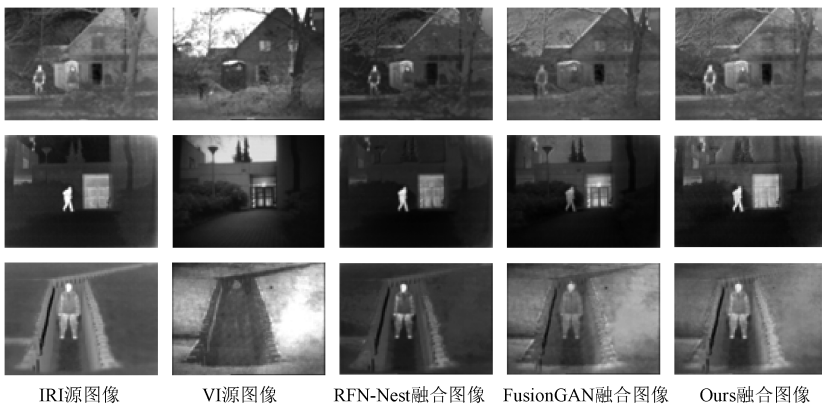


图 6 源图像和融合图像的定性比较

从图 6 可以看出:采用 RFN-Nest 得到的融合图像可视化效果一般,整体图像偏暗;而采用 FusionGAN 方法融合对红外图像中的高对比度

区域(如人及附近区域)保留得不充分;采用本文所提方法融合的图像对于两幅源图像各自的独特信息都有一定程度的保留,融合图像的可视化效

果较好。

选取 TNO 数据集,分别采用自适应稀疏表示(adaptive sparse representation, ASR)^[21]、曲波变化(curvelet transform, CVT)^[22]、基于制导滤波的融合(guided filtering based fusion, GFF)^[23]、梯度转移融合(gradient transfer fusion, GTF)^[24],以及前文提到的 FusionGAN、RFN-Nest、本文所提融合方法等对红外与可见光图像进行融合处理,对融合图像的信息熵和标准差进行定量评价,结果如表 1 所示。

表 1 融合图像的定量评价结果

融合方法	<i>H</i>	<i>S</i>
ASR	6.319 6	23.318 2
CVT	6.530 8	26.937 4
GFF	6.596 6	26.204 7
GTF	6.687 7	27.028 6
FusionGAN	7.061 8	40.761 8
RFN-Nest	6.841 3	35.270 4
Ours	7.130 1	43.286 2

从表 1 中可以看出,本文所提方法相比于其他融合方法,融合图像的信息熵和标准差均得到了一定的提升。

4 结论

本文提出了一种基于生成对抗网络框架、融合多尺度跳跃链接与局部鉴别器的图像融合网络。该网络生成器基于 U-Net,采用多尺度链接架构提取并融合图像多层特征,提高了图像深层特征信息的利用程度。同时鉴别器采用与传统鉴别器鉴别模式不同的局部鉴别器。在生成器与鉴别器对抗融合的过程中,生成器更加注重对源图像细节纹理的融合,通过对红外图像掩膜的提取利用,保留了红外图像的高对比度区域特征信息。实验结果表明,所提方法相较于其他经典的图像融合方法有更好的融合效果。

参考文献

[1] 唐霖峰,张浩,徐涵,等. 基于深度学习的图像融合方法综述[J]. 中国图象图形学报, 2023,

28(1): 3-36.
 [2] 李霖,王红梅,李辰凯. 红外与可见光图像深度学习融合方法综述[J]. 红外与激光工程, 2022, 51(12): 337-356.
 [3] 李晓军. 基于 NSST 的多模态医学图像融合算法研究[D]. 杭州: 浙江理工大学, 2019.
 [4] 杨桃,童涛,陆松岩,等. 基于多特征的红外与可见光图像融合[J]. 光学精密工程, 2014, 22(2): 489-496.
 [5] CHEN J, LI X J, LUO L B, et al. Infrared and visible image fusion based on target-enhanced multiscale transform decomposition [J]. Information Sciences, 2020, 508: 64-78.
 [6] YANG B, LI S T. Multifocus image fusion and restoration with sparse representation [J]. IEEE Transactions on Instrumentation and Measurement, 2009, 59(4): 884-892.
 [7] YANG Y, ZHANG Y M, HUANG S Y, et al. Infrared and visible image fusion using visual saliency sparse representation and detail injection model [J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 3011766.
 [8] MA J, ZHOU Z, WANG B, et al. Infrared and visible image fusion based on visual saliency map and weighted least square optimization [J]. Infrared Physics and Technology, 2017, 82: 8-17.
 [9] ZHANG H, YUAN J T, TIAN X, et al. GAN-FM: infrared and visible image fusion using GAN with full-scale skip connection and dual Markovian discriminators [J]. IEEE Transactions on Computational Imaging, 2021, 7: 1134-1147.
 [10] LI H, WU X J, KITTLER J. RFN-Nest: an end-to-end residual fusion network for infrared and visible images [J]. Information Fusion, 2021, 73: 72-86.
 [11] LI H, WU X J. DenseFuse: a fusion approach to infrared and visible images [J]. IEEE Transactions on Image Processing, 2018, 28(5): 2614-2623.
 [12] MA J Y, YU W, LIANG P W, et al. FusionGAN: a generative adversarial network for infrared and visible image fusion [J]. Information Fusion, 2019, 48: 11-26.
 [13] HINTON G E, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets [J]. Neural Computation, 2006, 18(7): 1527-1554.
 [14] GOODFELLOW I, POUGET-ABADIE J,

- MIRZA M, et al. Generative adversarial nets [C]// NIPS'14: Proceedings of the 27th International Conference on Neural Information Processing Systems-Volume 2. Cambridge, MA, USA: MIT Press, 2014: 2672-2680.
- [15] 朱辉, 秦品乐. 基于多尺度特征结构的 U-Net 肺结节检测算法[J]. 计算机工程, 2019, 45(4): 254-261.
- [16] FENG S L, ZHAO H M, SHI F, et al. CPFNet: context pyramid fusion network for medical image segmentation[J]. IEEE Transactions on Medical Imaging, 2020, 39(10): 3008-3018.
- [17] DENG Z J, HU X W, ZHU L, et al. R³net: recurrent residual refinement network for saliency detection[C]// IJCAI'18: Proceedings of the 27th International Joint Conference on Artificial Intelligence. Menlo Park, CA, USA: AAAI Press, 2018: 684-690.
- [18] ROBERTS J W, AARDT J V, AHMED F. Assessment of image fusion procedures using entropy, image quality, and multispectral classification [J]. Journal of Applied Remote Sensing, 2008, 2: 023522.
- [19] RAO Y J. In-fibre Bragg grating sensors [J]. Measurement Science and Technology, 1997, 8(4): 355-375.
- [20] TOET A. The TNO multiband image data collection[J]. Data in Brief, 2017, 15: 249-251.
- [21] YANG B, LI S T. Visual attention guided image fusion with sparse representation [J]. Optik, 2014, 125 (17) : 4881-4888.
- [22] NENCINI F, GARZELLI A, BARONTI S, et al. Remote sensing image fusion using the curvelet transform[J]. Information Fusion, 2007, 8(2): 143-156.
- [23] LI S, KANG X, HU J. Image fusion with guided filtering [J]. IEEE Transactions on Image Processing, 2013, 22(7): 2864-2875.
- [24] MA J Y, CHEN C, LI C, et al. Infrared and visible image fusion via gradient transfer and total variation minimization [J]. Information Fusion, 2016, 31: 100-109.

~~~~~  
 (上接第 8 页)

- [15] 赵晓娟, 贾焰, 李爱平, 等. 多源知识融合技术研究综述[J]. 云南大学学报(自然科学版), 2020, 42(3): 459-473.
- [16] 窦小强, 刘天雅, 张志政. 基于军事知识图谱的问答系统[C]// 第六届中国指挥控制大会. 北京: 中国指挥与控制学会, 2018: 537-541.
- [17] 张进, 徐宁骏, 赵伟光, 等. 基于智能客服技术的武器系统故障诊断方法[J]. 指挥控制与仿真, 2020, 42(4): 123-127.
- [18] 葛唯益, 王振宇, 王羽, 等. 主流知识图谱存储系统试验对比[J]. 指挥信息系统与技术, 2019, 10(5): 28-33, 75.
- [19] 梅发国, 戴大伟, 张冀. 基于知识图谱的战场目标关系融合技术[J]. 指挥信息系统与技术, 2017, 8(5): 81-86.
- [20] 王保魁, 吴琳, 胡晓峰, 等. 基于知识图谱的联合作战态势知识表示方法[J]. 系统仿真学报, 2019, 31(11): 2228-2237.
- [21] 张桂林, 王菁, 王惠娟, 等. 基于知识图谱的战场目标智能识别推理研究[C]// 第六届中国指挥控制大会. 北京: 中国指挥与控制学会, 2018: 326-330.
- [22] 张钺, 朱军, 苏航. 迈向第三代人工智能[J]. 中国科学: 信息科学, 2020, 50(9): 1281-1302.